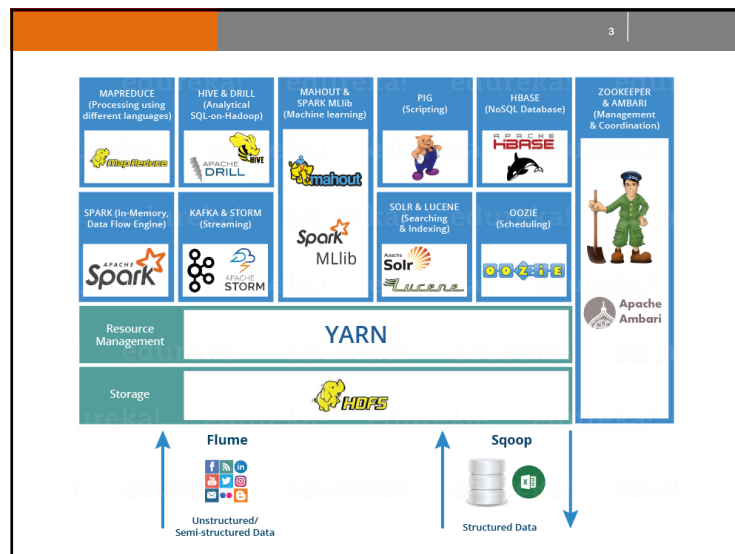
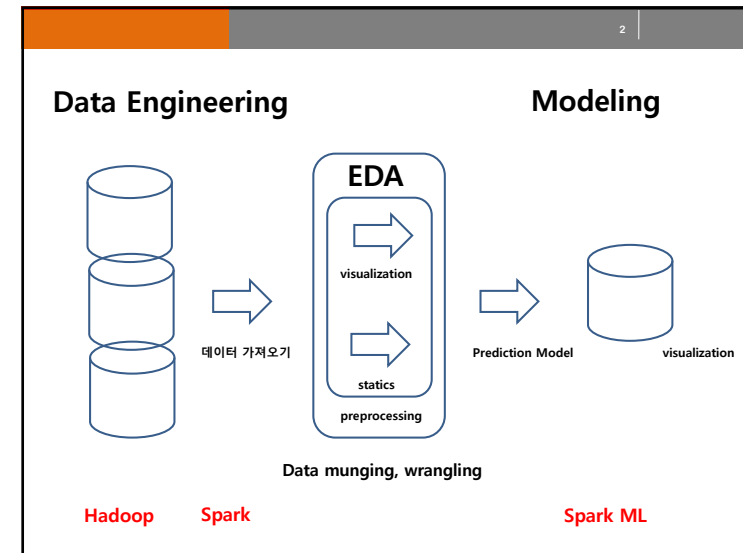


1

[표 1-3-1] 국내의 데이터 직무 구분 비교표

구분	한국데이터진흥원	한국정보화진흥원	고용노동부	노규성 박상학·박정혜	DATA FLOQ	SAS
데이터 설계	데이터 아키텍트				빅데이터 솔루션 아키텍트	빅데이터 아키텍트
데이터 개발	데이터 개발자	하둡/NoSQL /MapReduce /엔지니어	DB 엔지니어	빅데이터 관리 및 기술(지원)	빅데이터 엔지니어	빅데이터 개발자
데이터 운영관리	데이터 엔지니어					
데이터 활용	데이터 분석				빅데이터 분석가/빅데이터 사이언티스트	빅데이터 분석가/빅데이터 사이언티스트
	데이터 시각화	빅데이터 분석가/데이터 사이언티스트		빅데이터 분석	빅데이터 비주얼라이저	빅데이터 디자이너
	데이터 연구		빅데이터 분석		빅데이터 연구원	데이터 사이언티스트
데이터 관리	데이터 사이언티스트/DBA			빅데이터 활용/데이터 관리 및 기술(지원)	총괄책임자/빅데이터 매니저/빅데이터 사이언티스트	프로젝트 매니저/데이터 사이언티스트/빅데이터 관리자
데이터 판매	데이터 마케터	빅데이터 기술영업/마케터	-	-	-	-
데이터 컨설팅	데이터 컨설턴트	빅데이터 컨설턴트	-	빅데이터 기획	빅데이터 컨설턴트	-

출처: "데이터 진문인력 양성 방안 연구 결과 보고서", 한국데이터진흥원, 2016



4

데이터분석

- Numerical, Graphical
- Trial and Error
- Ad hoc 일반화할 수 없는 해결책
- No Free Lunch 일반화, 보편적 기법 < 특화된 기법

D. H. Wolpert. The supervised learning no-free-lunch theorems. In Soft Computing and Industry, 25–42p. Springer, 2002.

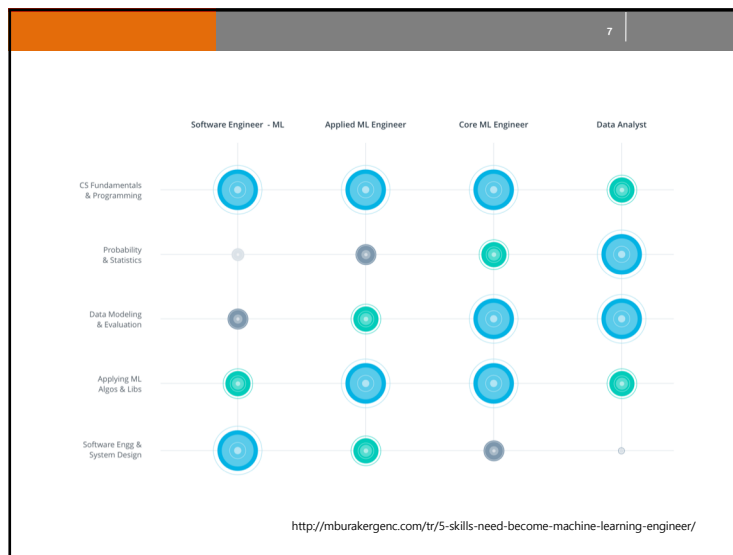
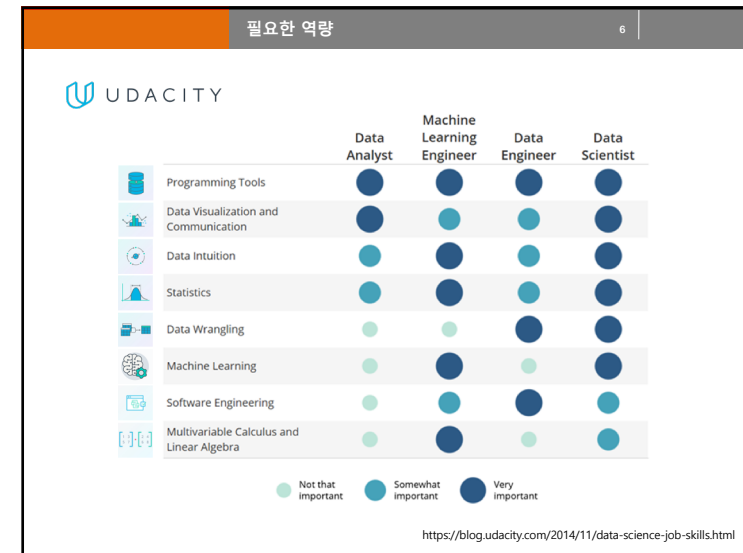
5

다양한 시도, 새로운 기법
GUI 기반 툴 < **프로그래밍**

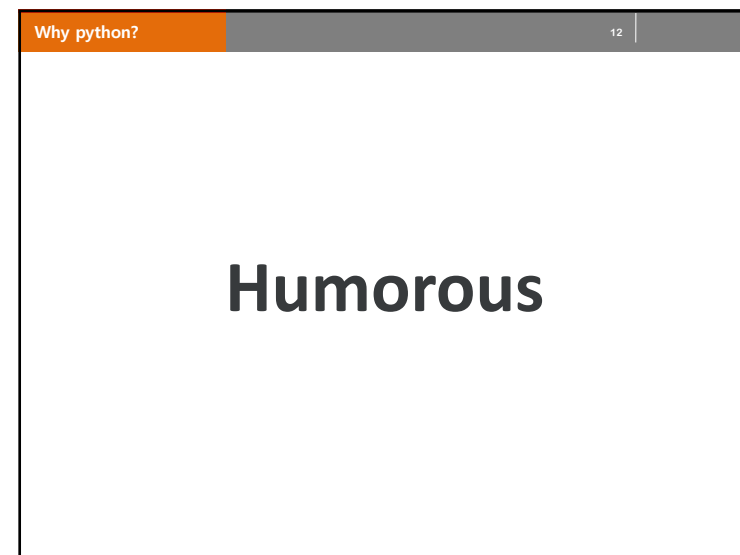
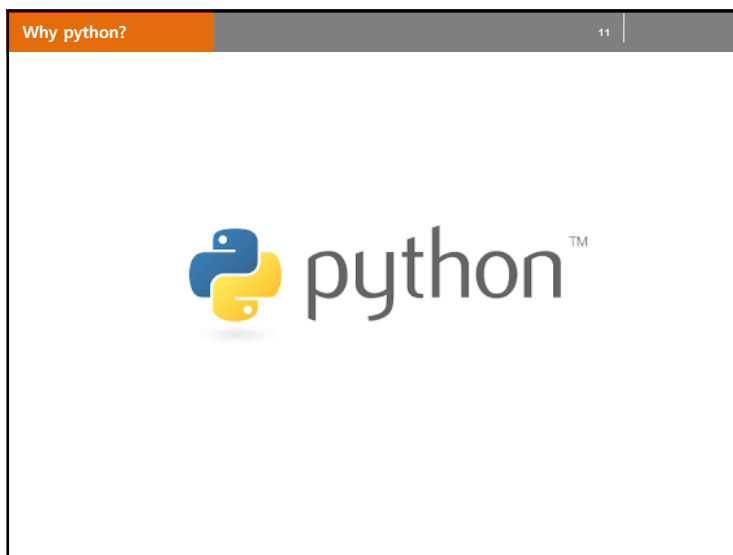
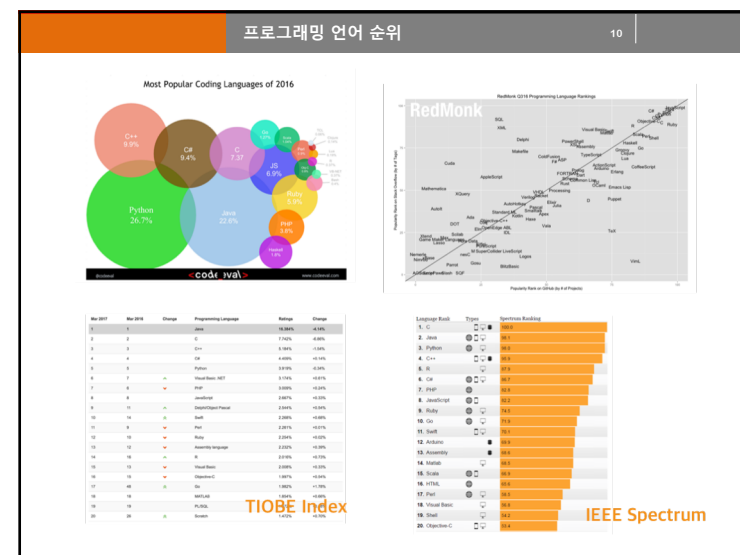
시행착오
Compiler < **Interpreter**
-> Jupyter Notebook

느린 속도
-> 선형대수 기반 (N차원 벡터)
-> C언어 기반 라이브러리 (Numpy)


데이터 분석 관련 패키지 통합/관리
-> Anaconda (Numpy, Scipy, Scikit-Learn, Pandas, Keras, ...)



Python 기반의
기계학습 Framework를
활용한 Data 분석




Why python? 13



<https://gvanrossum.github.io/>
<https://www.youtube.com/watch?v=jiu0IYQIPqE>

Why python? 14

Life is too short, You need



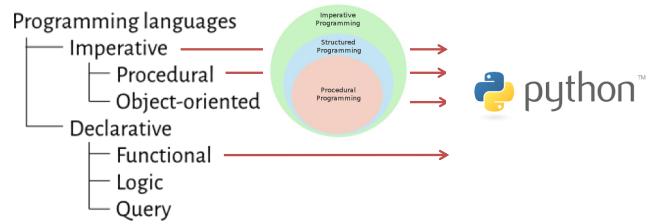
Why python? 15

생산성

**Effective
Efficient**

Why python? 1. Multi Paradigm 16

Multi Paradigm



Why python? 2. Glue Language 17

Glue Language



python™
[Language Reference](#)

- CPython (c.f : cython)
 - De facto
 - Python Software Foundation 관리
- 대체, 플랫폼 특정 구현체
 - PyPy, Stackless
 - IronPython, PythonNet, Jython
 - Skulpt, Brython
 - MicroPython

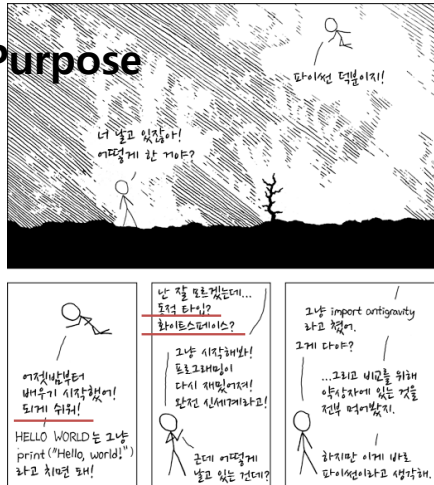
Why python? 3. 다양한 라이브러리 & 도구들 18

Library & Tool

- 다양한 종류의 수많은 **standard library** 기본 탑재
플랫폼에 상관없음
- 다양한 종류의 수많은 **open-source libraries**
the **Python Package Index**: <https://pypi.python.org>
pip
the de facto
default package manager

Why python? 4. General Purpose and ... 19

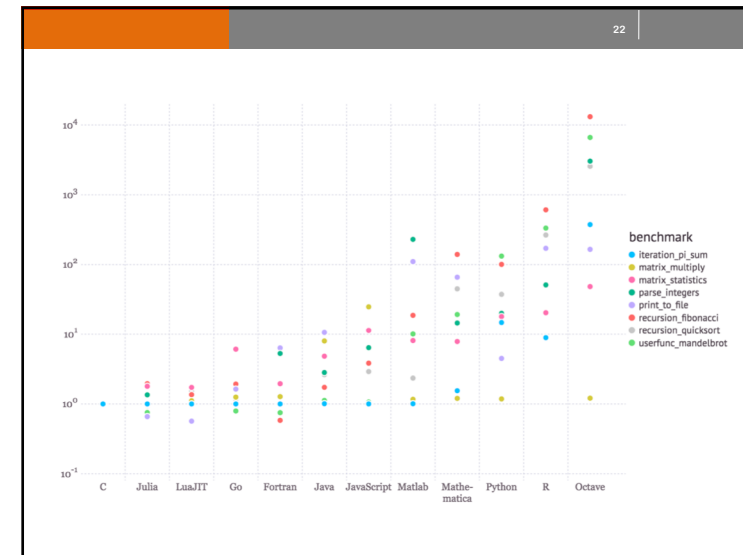
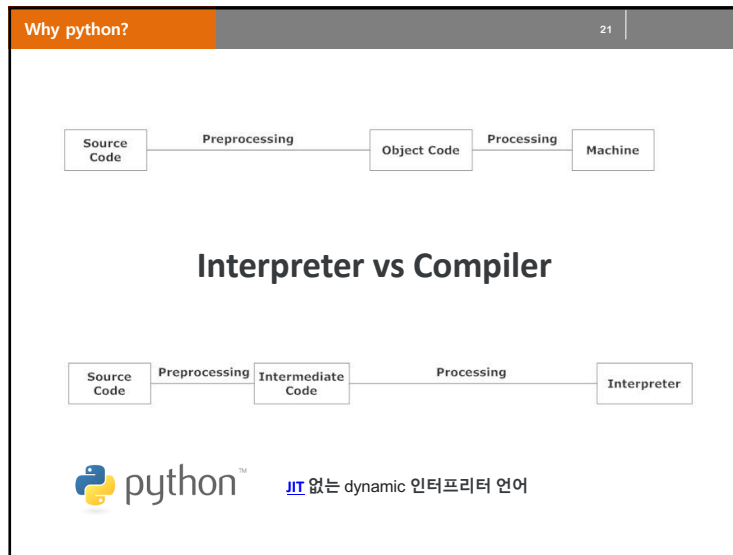
General Purpose



C.f)
Domain-specific

Why python? 20

Interpreter Language



Python 기반의 기계학습 **Framework**를 활용한 Data 분석

24

■ Jupyter Notebook

- 데이터 분석
 - 여러가지 알고리즘과 파라미터를 바꿔가면서 분석 결과를 반복해서 확인 (Graph나 Table 형태 포함)
 - 컴파일 언어, Console 환경에서는 불편
- Notebook 환경
 - 코드작성 뿐만 아니라 일반문서 및 보고서 그리고 데모를 위한 프레젠테이션을 작성-실행-결과확인-코드수정을 반복 (REPL) 하면서 원하는 결과를 만들어 낼수있는 작업환경
 - Matlab, Mathematica
 - Rstudio
 - Web
 - » Zeppline, Jupyter Notebook
 - » 협업 기능
 - Notebook 환경이 De Facto

■ Jupyter Hub

■ Jupyter Lab

